

5 **MAINTAINING EYE-CONTACT IN TELECONFERENCING**
 USING STRUCTURED LIGHT

CROSS-REFERENCE TO RELATED APPLICATIONS

10 This application is a Continuation Application of U.S. Patent Application Ser.
 No. 10/121,562, filed April 11, 2002, and Provisional Patent Application Ser.
 No. 60/283,158, filed April 11, 2001, which are incorporated herein in its
 entirety by this reference made thereto.

15 **BACKGROUND OF THE INVENTION**

TECHNICAL FIELD

 The invention relates to teleconferencing. In particular, the invention relates to
20 methods and systems that permit the appearance of eye-contact to be
 maintained between participants in a teleconference.

DESCRIPTION OF THE PRIOR ART

25 A primary concern with video teleconferencing systems is the frequent lack of
 eye-contact between participants. In the most common configuration, each
 participant uses a computer monitor on which an image of the remote
 participant is displayed, while a camera mounted above the monitor captures
 an image of the local participant for display on the monitor of the remote
30 participant. Because participants frequently look at either at the image of the
 remote participant or elsewhere on the display, rather than directly at the
 video camera, there is the appearance that the participants are not looking at
 one another. This results in an unsatisfactory user experience.

5 Prior art solutions to the eye-contact problem have incorporated half-silvered, partially transmissive and partially reflective mirrors, or beamsplitters. These solutions have typically incorporated a beamsplitter placed in front of a computer display at a 45 degree angle. In one typical configuration, a video camera, located behind the beamsplitter, captures the image of the local
10 participant through the beamsplitter. The local participant views an image of the remote participant on the display as reflected by the beamsplitter.

In devices incorporating a conventional CRT, the resulting device is both bulky and physically cumbersome. In cases involving an upward facing display, the display is viewable both directly and as reflected by the
15 beamsplitter, greatly distracting the local participant. To alleviate this problem, prior solutions, including those described in U.S. patents 5,117,285 and 5,612,734 have introduced complicated systems involving polarizers or micro-louvers to obstruct a direct view of the upward facing display by the local participant. In all cases, the image of the remote participant appears
20 recessed within the housing holding the display, beamsplitter, and video camera. The resulting distant appearance of the remote participant greatly diminishes the sense of intimacy sought during videoconferencing.

Another set of prior art attempts seeks to alleviate this problem through the use of computational algorithms that manipulate the transmitted or received
25 video image. For example, U.S. patent 5,500,671 describes a system that addresses the eye-contact problem by creating an intermediate three-dimensional model of the participant based on images captured by two video cameras on either side of the local display. Using this model, the system repositions artificially generated eyes at an appropriate position within the

5 image of the local participant transmitted to the remote participant. The
resulting image, with artificially generated eyes and a slight but frequent
mismatch between the position of the eyes relative to the head and body of
the participant, is unnatural in appearance. Furthermore, the creation of an
intermediate three-dimensional model is complex and time-consuming,
10 making it difficult to implement in practice.

A further weakness of these and other similar approaches is an inability to
handle all possible participant postures and movements. More robust
algorithms are possible and several have been proposed, but these
approaches are more computationally complex, and cannot be executed
15 rapidly enough on current microprocessors to allow for real time processing of
high resolution video images. Finally, many of these approaches require that
the remote communicant own and operate the same videoconferencing
device. This presents a significant obstacle to introduction and widespread
adoption of the device.

20 What is needed is a device that incorporates at once all of the beneficial
features achieved by the prior art, while addressing the aforementioned
deficiencies. First and foremost, the device must offer eye-contact in a robust
manner, operating effectively across the full range of local participant head
positions and gaze directions. It must provide a natural view of the remote
25 participant for the local participant. It must be aesthetically pleasing and
easily operated by a typical user. The underlying algorithm must be
computationally simple enough to be conducted in real time on high frame
rate, high resolution video. Finally, the device should require little if any

5 additional videoconferencing equipment beyond that found in a typical existing videoconferencing setup.

SUMMARY OF THE INVENTION

10

The invention comprises a structured light projector, a video camera, and an image processor, for achieving perspective corrected images that enhance eye-contact during teleconferencing. A structured light projector is offset in one direction from the monitor center, and illuminates a local participant with a structured light pattern. The image of the local participant, illuminated by both ambient and structured light, is captured by the video camera, also offset from the monitor center, preferably in the direction opposite the structured light projector. By considering the distortion of the structured light observed from the position of the video camera and the position of the structured light projector and video camera relative to the monitor center, an image processor creates an image of the local participant as viewed from a perspective that, when viewed by the remote participant, provides a sense of eye contact with the local participant.

25

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 is a block schematic diagram which shows a preferred embodiment of the invention;

30 Figure 2 shows a local participant illuminated by ambient and structured light according to the invention;

5

Figure 3 shows the result of a line detection operation according to the invention;

Figure 4 shows the result of a directional convolution applied to the results
10 shown in Figure 3, where 1, -1, and 0 are represented by white, black, and gray pixels, respectively;

Figure 5 shows the head outline of the local participant used for the image of Figure 2;

15

Figure 6 shows the filtered result of a warping calculation according to the invention; and

Figure 7 shows a final image produced according to the invention.
20

DETAILED DESCRIPTION OF THE INVENTION

Physical Description of the Invention

The presently preferred embodiment of the invention, as shown in block
25 schematic form in Figure 1, comprises three primary components:

- a structured light projector;
- a video camera; and
- an image processor.

5 The structured light projector 40 provides a source of structured light. In the preferred embodiment, the structured light projector projects a pattern of infrared light, so that the structured light is not visible to either the local or remote participant. The structured light pattern preferably comprises a series of horizontal lines. In the discussion of and figures for the preferred
10 embodiment that follow, the pattern of infrared light is illustrated as white lines on a black field -- white corresponding to full infrared illumination and black corresponding to no infrared illumination. The width of each horizontal line is approximately equal to the spacing between adjacent lines. Satisfactory results may be achieved with a pattern comprising approximately twenty such
15 lines.

Such a projector can be constructed by replacing the existing bulb in a standard slide projector with an infrared light source, preferably a high output infrared light emitting diode. In this case, the desired structured light pattern may be reproduced on a slide inserted into the projector. Alternatively, a
20 structured light projector may be obtained through the modification of a video projector, for example an LCD video projector.

Finally, several commercial products for producing structured light are well known in the art. For example, the invention may be practiced with a Stocker Yale Lasiris structured light laser.

25 The video camera 30 may be any known device for capturing images of the local participant 10 that is also capable of capturing the structured light pattern. To ensure that the Infrared structured light is undetectable by the remote participant, it may be desirable to use a camera having a separate channel for infrared image content. Such a camera collects and transmits

5 the infrared image content in an infrared channel, I, alongside the standard R, G, and B channels. By eliminating the I channel from the transmission to the remote participant, the structured light is removed from the image of the local participant.

More commonly among video cameras offering infrared sensitivity, the
10 infrared content of the image is mapped into the RGB channels transmitted to the remote camera. Accordingly, any infrared structured light captured by the camera would be displayed on the remote monitor within the human eye's sensitivity range. In this case, the structured light may be removed through the use of timing circuitry. This circuitry coordinates the structured light
15 projector and video camera. Structured light is presented only periodically and for a short duration, for example less than the duration of a single frame of the video camera. The frames captured during structured light illumination are not transmitted to the remote location. In place of such frames, the previous frame may be repeated. Past experience converting film between
20 formats with different framing rates has shown that the human eye cannot detect the occasional repetition of a single frame. Alternatively, the structured light may be presented in the vertical blanking interval of a first video camera, and captured by a second video camera.

The image processor 50 implements the inventive technique, which is
25 discussed in greater detail below. The image processor is in communication with the video camera 30, and in some embodiments, the image processor is also in communication with the structured light projector 40.

A monitor 20 allows the local participant to view the remote participant.

Operation of the Invention

General Operation

5 The structured light projector 40 is preferably offset in one direction from the monitor 20 center, and illuminates the local participant 10 with a structured light pattern, in the preferred embodiment a series of substantially parallel lines. The lines are preferably oriented substantially perpendicular to the displacement of the structured light projector from the monitor center. An
10 image of the local participant, illuminated by both ambient and structured light, is captured by a video camera 30 that is offset from the monitor center, preferably in a direction opposite that of the structured light projector.

The pattern of structured light projected onto the local participant appears as substantially straight, evenly spaced lines from, and only from, the perspective
15 of the structured light projector. From all other perspectives, including that of the video camera, the lines of structured light are distorted as they traverse the physical features of the local participant.

By considering the particular distortion observed from the position of the video camera, the image processor is capable of producing an image of the local
20 participant as viewed from perspectives other than that of the video camera. This is accomplished by first isolating the lines of structured light, and then calculating the amount of warping needed to restore the lines to a straight configuration. Performing the warping determined in this manner yields an image of the local participant as viewed from the position of the structured

light projector. By performing only a fraction of the warping determined in this manner, the image processor can obtain an image of the local participant as viewed from a point along the line between the video camera and structured light projector. In particular it is possible to obtain an image of the local participant as seen from the monitor center. Such a view point, when displayed on the monitor of the remote participant, provides the remote participant with a sense of eye contact with the local participant.

It should be noted that optimal eye contact is achieved by providing an image of the local participant as seen from the location of the remote participant's eye on the local display. Typically, this point is very near the monitor center. However, in some embodiments of the invention, an adjustment may be made to more accurately track the position of the remote participants eyes on the local display and adjust the amount of warping performed accordingly

A more detailed description of the process performed by the image processor is provided below.

Line Detection

Figure 2 shows the local participant illuminated by ambient and structured light according to the invention. The image processor begins the process of determining the requisite warping by isolating as precisely as possible the structured light lines from the image. In the preferred embodiment, the lines are detected by thresholding the results of a high pass convolution filtering operation. This can be summarized as

$$L = T_i(H * G_1(P))$$

where P is the original image, G_1 is, for example, an 11 x 11 elliptical Gaussian filter having a horizontal standard deviation of two pixels and a vertical standard deviation of one pixel. H is a high pass filter, and T is the numerical value of the threshold operator. For example,

$$H = \begin{bmatrix} 0 & -1 & -2 & -1 & 0 \\ -1 & 1 & 2 & 1 & -1 \\ -1 & 1 & 4 & 1 & -1 \\ -1 & 1 & 2 & 1 & -1 \\ 0 & -1 & -2 & -1 & 0 \end{bmatrix}$$

Alternatively, this operation can be performed on the difference between images obtained with and without structured light illumination. Specifically,

$$L = T_l(H * G_1(P - P'))$$

where P is the original image with structured light and P' is the original image without structured light.

Figure 3 shows the result of the line detection operation. The image processor has succeeded in isolating the structured light lines, defining them more clearly than in Figure 2. To determine the warping needed to return these lines to a straight configuration, the image processor first convolves the modified image, L , of Figure 3 with two directional operators

$$S = D_{se} * (D_{ne} * L)$$

Here,

$$D_{se} = \begin{bmatrix} 0 & -1 & 0 \\ 1 & 0 & -1 \\ 0 & 1 & 0 \end{bmatrix}, \quad D_{ne} = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & -1 \\ 0 & -1 & 0 \end{bmatrix}$$

The result of this directional calculation is a tri-valued image. Those pixels within the image through which a southeasterly line passes are valued 1, pixels through which a northeasterly line passes are valued -1, and all other pixels are zero-valued.

- 5 Figure 4 shows the result of the directional convolution applied to Figure 3, where pixel values of 1, -1, and 0 are represented by white, black, and gray pixels, respectively. The image processor then obtains a measure of the required warping by integrating this tri-valued image along lines parallel to the undistorted lines of structured light. For simplicity, the required warping may
10 be determined and performed only within a region coincident with the head outline of the local participant, and the background is left unaltered.

- Figure 5 shows the head outline of the local participant for the image of Figure 2. Accordingly, the integration is performed along lines parallel to the undeformed lines of structured light, with the limits of integration defined by
15 the white region shown in Figure 5. For horizontal lines of structured light, the required upward warping of each point is given by the sum of all pixel values left of the point but within the white outline of Figure 5. Specifically, for lines of structured light aligned with a horizontal x-axis,

$$W(x, y) = \int_{x_0(y)}^x S(x, y),$$

- 20 where $x_0(y)$ is a lower limit of integration determined by the left edge of the region defined in Figure 5.

Figure 6 shows the filtered result of the required warping calculation. The image processor uses a second filter, for example a circular Gaussian filter, to smooth the results to those shown. The results of this calculation indicate the

warping required to return the structured lines to an undistorted configuration. Applying this warping to the image L provides an image of the local participant as seen from the viewpoint of the structured light projector.

- 5 To achieve an image of the local participant as viewed from the monitor display center, a fraction, approximately half, of this warping is performed. The precise fraction is preferably determined by the ratio of the camera to structured light projector distance and the camera to monitor center distance.

- 10 The may be accomplished with the "meshwarp" image warping routine developed by Douglas Smythe and Industrial Light and Magic. [A Simplified Approach to Image Processing – Classical and Modern Techniques in C, Randy Crane, Prentice Hall PTR, 1997, pp 223-230]. This algorithm constructs a new image given an initial image and a set of displacements for each pixel. Alternatively, an image warp based on bilinear interpolation or field based warping may be employed.

- 15 Regardless of the specific routine used, the result of this process is an image showing an estimate of the local participant as seen from the display center. This image is shown in Figure 7.

- 20 Although the invention is described herein with reference to the preferred embodiment, one skilled in the art will readily appreciate that other applications may be substituted for those set forth herein without departing from the spirit and scope of the present invention. Accordingly, the invention should only be limited by the Claims included below.